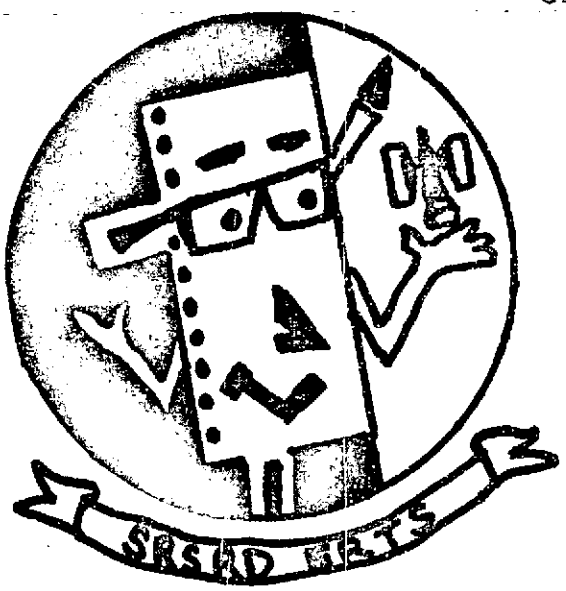
Type II Progress Report

June 20, 1973 - December 19, 1973

Crop Identification & Acreage Measurement Utilizing ERTS Imagery 013



Principle Investigator

Donald H. Von Steen AG 328

TABLE OF CONTENTS

GROUND DATA [1/]

Ground Survey Observations

The initial ground observations were obtained in June of 1972. Because of the uncertainty of ERTS's launch, the update observations did not begin until August of 1972. The ground observations have been summarized by each study area. The coefficients of variation for major crops ranged from about 10 to 40 percent with some of the lesser prominent crops ranging from about 30 to 70 percent. It is expected that ERTS imagery interpretation results will be highly correlated to the ground observations, and if they are substantial gains in precision, can be obtained for areas similar to the study areas by using ERTS imagery in the estimation procedure.

Cost Analysis

The average cost of collecting the initial ground observations was found to be about $32.50 per area segment. To collect the update information, it cost on the average about $13.70 per area segment. The difference between the two cost figures represents the additional costs required to locate the June segment operators, secure crop intentions, secure livestock data and farm labor data. The ERTS Update fieldwork only included locating the segments and recording the crops present and their condition. the operators were not contacted unless the enumerator could not view the fields from the road.

Aircraft costs computed from cost estimates provided by Mr. Bernie Nolan of NASA, indicate it would cost about $60 per area segment. This figure only includes the cost of acquisition. The interpretation and summarization of aircraft data has not been determined.

The only cost of obtaining ERTS data that we have been able to obtain is the cost of purchasing the CCT's from Sioux Falls, which is $160 per ERTS scene. Our study areas require about three scenes to obtain complete coverage which gives a cost of about $9 per segment. It is our understanding that the $160 does not include the cost of launching ERTS or the cost of maintaining the satellite in orbit. Because of this, the costs of acquiring data by the three collection methods are not exactly comparable.

1/ Summarized from the Progress Report dated August 20 - October 19.

MICRODENSITOMETER

General

The microdensitometer was installed on August 1, 1973 by Richard Ulinski, a representative of the Perkin-Elmer Corporation. During installation, it was discovered that the high speed reader punch was not included. The punch has been received by us and is expected to be installed the week of Janurry 6, 1974.

The scanning apertures and microscope objectives we need to perform our work have been ordered. Delivery is expected to be the week of January 6. As soon as they are installed, we will begin scanning ERTS aerial photography.

During processing of some service work for Don Klinglesmith and Robert Mercer, the position measurement system developed a problem, and the microdensitometer stage would run away in one axis. Two new circut boards are being sent to correct the trouble.

Support Service

Supporting services were provided to Robert D. Mercer from Dudley Observatory, for Low Brightness Image Data Analysis (NAS 9-12557) of photography returned from apollo lunar Orbital Missions. This information showed digitization on the PDS microdensitometer of 35mm original data frame, photometric calibrations and lens brightness transfer function photography. This digitized information will be further processed through the NASA VICAR Program at GSFC's Computer facility to provide the investigator with absolute photometry. Isophote maps and mosaic scenes of such astronomical phenomena as the Zodiacla light, linear calibartion regions, galactic sources and genenschein. The quality of data processing is inhanced for both projects with an estimated saving to NASA of at least $5,000 by sharing in the use of the PDS microdensitometer for this work has helped to uncover speicification deficiencies prior to final equipment acceptance by USDA and at the same time have saved considerable travel and vendor digitization service expenses for the Low Brightnes Image Data Analysis work.

SOFTWARE SUPPORT

Convert Microdensitometer Data into a SAS Compatable Format

The PDSCMS program is now running for all features, providing there are no serious errors. The fatal error detection and recovery logic has yet to be thoroughly tested. A complete program manual is included as Appendix A. However, the major program features are described here.

A program will produce SAS type observations from multiple scans of the same scene. In addition, certain housekeeping functions are performed to facilitate processing the scan picture in SAS. Some of these are:

(1) Verify that the pictures are correct and compatible.

(2) Remove scan raster.

(3) Affix user defined symbolic identifiers.

(4) Compute and assign an x and y coordinate position for each pixel.

Convert Microdensitometer Data into Penn State Format

The development of this program has been slowed because the programmer/analyst assigned to the project left us for another agency. There were no salvageable results. Our agency has selected another programmer/analyst, George Howse, to work part-time. We expect him to begin near the end of January.

Penn State Classifier: Version 1

Version 1 of the Pennsylvania State Classifier has been up and running error free since June '73. The package was installed over a period of 6 months without working full-time on it. Most items were running within the first 3 months, and the remaining 3 month period was spent correcting installation dependent problems as they occured. Our impressions follow:

1). Dr. Borden and associates deserve a great deal of credit for providing a simplified classified classification system. The basic subset file is common to all processing programs, and tends to unify the package. This seems to me, a Herculean effort when they have no central program library and some semi-independent programming.

2). The system uses a unified control card language to control the program operations. Each control card is identified by a keyword followed by options. This feature makes it much easier to learn and use the system.

3). The package is mostly written in FORTAN, which means that it can be more easily moved to another installation, or different computers than a package that uses alot of assembly code.

4). The required assembly coded software is in the form of installation dependent routines. These are provided by the computation center to make work easier for their users, but they are not necessarily available or compatible everywhere. Some of these routines such as the date function should be generally available. The input-output routine, FASTIO, was provided to us by the Penn State Computer Center and worked perfectly. The specialized REREAD routine was supplied but could not be made to work. A replacement routine @INCORE is used instead.

5). The program was installed here in a centralized library that contains only one copy of any program or subroutine. We found that there were 2 versions fo CLASS, 2 versions of SCALE, 3 versions of GETLIN, 2 versions of RECTIF, and 2 versions of OPEN. These different versions performed essentially the same task but slightly different. In order to include these mirror copies, we had to hunt them down in the original program tape, delete off everything else, and remove the subroutine so it could be placed in the program library.

In my opinion, mirror copies are a serious programming defect. Whenever a functional problem has occured, fixes are required in 2 or 3 places, and each fix is fresh code. Also each problem may mean a new copy of an already existing subroutine. These mirror copies have caused us a great deal of troble and will continue to hamper development of this system until they are eliminated.

6). An attempt was made to operate the programs under TSO. This did not turn out very successful because (1) the TSO response time was too slow to justify the waiting time, and (2) the programs tended to print too much material for a typewriter terminal.

7). We were not able to have the programs create and save files for subsequent runs. At Penn State, they use something called a BAT file to pass the Statistics to the classifier program. This is an installation procedure and is transparent to the user. We attempted to use a partitioned data set, but the IBM support Software failed. Thus, we were obliged to normally transfer data between programs, and inconvenience.

8). The system may use more CPU time than necessary. A special mappint program was developed inhouse, using standard IBM FORTRAN that ran from 1 1/2 to 2 times faster than the Penn State NMAP program. The comparisons may not be completely valid because the NMAP program has a great deal more flexibility. The inhouse program is limited to a single band and must process every point in a given block. The NMAP program can perform multiband maps, and every n'th point and line in a given block.

9). The programming package does not have a quadratic discriminate function. This is considered a serious defect and has resulted in restricted usage for processing ERTS data.

10). The ACLASS discriminate function provided, performs a kind of normalization that is expected to produce good or better results when processing aircraft data. The normalization attempts to minimize the effect of sun angle. This feature ought to be very useful for processing the microdensitometer data.

Penn State Classifier: Version II

This program has been separated into three groups. Group I has been split into separate decks. Version I source decks have been removed from the program library, and version II decks will be processed and moved into the program library over the next month.

SEGMENT LOCATION

A random selection of land segments for ground enumeration was made. Random selection has a twd-fold advantage. (1) The data is a representative, and (2) is the ability to expand the ground data into an area estimate. Random selection does pose problems of finding this location in ERTS scenes.

The locating of segments and fields within segments was a big task. First, the segments are located and drawn on county highway maps. In addition, the segments are also located on large 24x24 and smaller 9x9 ASCS aerial photographs and individual fields are drawn in and the crop identified.

Special color IR aerial photography was taken over selected segments during the growing season. These were supposed to be made on or about the same date as the satellite went over. Segments were located within these flightlines by comparing gross landmarks and highways with the county maps.

Large blow-ups, 38x38, were made from selected ERTS images. If there was aircraft coverage for that area, the flightlines were drawn in. Segments within the flightlines were relatively easy to locate because of the correspondence with aerial photography. Segments outside flightlines were more difficult and had to be carefully measured from corresponding land features on the highway map and the ERTS photo.

Finally, and most difficult the segments locations were found on grey scale printouts from the ERTS MSS tapes. Generally, only gross features were visible in the computer printout and most segments were found by measuring from known features. After the segments were found, the field boundaries were drawn in using the ASCS photography as a guide, and color IR aerial photography when available.

The segments and fields must be precisely located on the MSS tapes in order for the computer to identify the crops in the ERTS image. A detailed write-up of the location procedure is in Appendix B.

A SUMMARY OF RESULTS

Analysis of the Idaho and Missouri test sites were performed during the reporting period. Results of temporal overlays, equal and unequal prior probabilities, independent test data are discussed. The amount of improvement that each technique contributed is summarized below:

1. The results in Missouri where temporal overlays were made, show that temporal information improved the overall classification by 10%.

2. The dates were not optimum that were overlaid.

3. Data analysis in both Missouri and Idaho indicates that the use of prior probabilities improves the overall classification rates by at least 10% overusing the assumption that the crops are all equally likely.

4. Using both procedures together indicates that overall performance can be improved by 20% over one date and equal prior probabilities.

5. Idaho data has banding problems that may have caused serious problems in the crop classification.

6. The twelve crop types in Idaho seem to be quite similar spectrally, and hence, classification is quite difficult.

7. ERTS may not contain enough information to have perfect classification, but the data may still be useful for making crop acreage estimates.

8. Remotely sensed data could be used with a regression estimator if there is a correlation between ground data and classification results.

9. Remotely sensed data could be used with a double sampling model if 8 above holds.

10. Also, a mixture problem approach is presented that may have potential.

The results of our analysis during this period were done on the Missouri and Idaho test sites. This analysis was done at Purdue on the LARS computer. Similar analyses will be done for test sites in Kansas, and South Dakota.

Each test site covers approximately 10,000 square miles. There are many segments, each about a mile square at each test site. These segments constitute a random sample of the test site. The ground enumeration or ground truth information is taken from these segments, and the data is used for training and testing. There are 52 segments at the Missouri test site and 44 segments at the Idaho test site. A circle drawn around all segments at a given test site would enclose about 10,000 square miles.

The results are presented in a classification matrix. Missouri will be presented first. Table 1 is an example of a classification matrix using quadratic discriminant functions with equal prior probabilities. That is, we have assumed that the probability of corn is the same as the probability of cotton, and so forth. In this table, the classification was made using the whole data set as both training and testing data. We also used data from 3 ERTS overflights. That is, data that has been temporary overlaid by the people at LARS. The left column contains cotton, corn, soybeans, grass, winter wheat, and odd. The next column gives the number of sample values in each of the crop classes. For cotton we have 927 pixels. Notice that 689 of those pixels are classified correctly. That is, 74.3%. The remainder were misclassified as follows: 21 of those 927 pixels were classified as corn, 83 as soybeans, 36 as grass, 61 as winter wheat, and 37 as an odd group. It should be pointed out that winter wheat had been harvested at this time and probably should have been included in the odd group. The overall performance in this table was 58.4%, that means we summed the correctly classified pixels (1295), and divided by the total number of pixels (2217). The thing to be stressed in this table is that equal prior probabilities were used. This assumption is obviously not valid, but is frequently used because of lack of information. In the second table, the analysis is the same except that now we have used unequal prior probabilities. These prior probabilities may be derived from last year's census data or an earlier survey in the same year. Our prior probabilities came from an earlier survey, the June 1972 Enumerative Survey, which was updated to the time of the first ERTS date. If we compare the two tables, one can see two facts: 1).Overall classification is much better in Table 2 than in Table 1, 2) The total number of pixels in the columns for each crop is now very close to the actual number of pixels. For example, from Table 2, the total number of pixels that were classified as cotton is 906. That number is considerably closer to the 927 which is the actual number of pixels present. Corn, likewise, has a total number of pixels, 43 and that is rather close to 58. For soybeans, we come out with a total of 866, that is very close to 852. The grass group or crop came out to have 277 pixels as compared to 240 actual pixels in this crop. Winter wheat had 27, compared to 85 actual pixels, and the odd group had 98 compared to 55. The winter wheat and odd classification indicate the importance of correct timing because as pointed out earlier, most winter wheat was stuble at this time. The overall performance was 70.5 which is a significant improvement in the overall performance in Table 1. Further, the statistical properties of estimates made on this basis are better since normality for the data set and the prior probabilities are correct, we obtain unbiased estimates of the crop categories.

Table 1--Classification matrix of quadratic discriminant functions with prior probabilities using data from 3 overflights[1/]

Group	No. of sample points	Percent correct	Number of samples classified into					
			Cotton	Corn	Soybeans	Grass	Winter wheat	Odd
Cotton	927	74.3	689	21	83	36	61	37
Corn	58	58.6	4	34	3	10	5	2
Soybeans	852	39.7	101	29	338	137	199	28
Grass	240	57.1	34	22	22	137	20	5
Winter wheat	85	69.4	5	2	6	7	59	6
Odd	55	69.1	9	3	1	2	2	38
Totals	2217		842	131	453	329	346	116

Overall performance 58.4

1/ August 26, 1972, MSS bands 4,5,7.
September 14, 1972, MSS bands 5,7.
October 2, 1972, MSS bands 4,5,6,7.

Table 2--Classification matrix of quadratic discriminant functions with unequal prior probabilities using data from 3 overflights[1/]

Group	No. of sample points	Percent correct	Number of samples classified into					
			Cotton	Corn	Soybean	Grass	Winter wheat	Odd
Cotton	927	79.7	739	2	137	26	0	23
Corn	38	44.8	9	26	7	14	0	2
Soybeans	852	71.8	99	12	612	96	8	25
Grass	240	53.3	42	1	66	128	0	2
Winter wheat	85	22.4	9	1	40	10	19	6
Odd	55	70.9	8	1	4	3	0	39
Totals	2217		906	43	866	277	27	98

Overall performance 70.5

1/ August 26, 1972, MSS bands 4,5,7.
September 14, 1972, MSS bands 5,7.
October 2, 1972, MSS bands 4,5,6,7.

Most classification reported by other researchers has not been based on the use of these prior probabilities, while the overall error rate reported here is higher than reported by some researchers. This study was based on a statistical sampling of the entire land area in the study areas and not purposely selected study sites. Consequently, the improvement in the classification using this technique is important. Classification is improved by about 10 percent, although this is a function of how unequal the sets are. Secondly, we would like to point out that when a data set is based on a probability sample, the user (SRS) is able to estimate these prior probabilities and take advantage of this procedure.

The next table shows results of the per field classifier. Point classifiers were used in the previous tables. Each pixel in a field can be assigned to any of the six groups in a point classifier system. In the per field classifier all pixels in the field are assigned to the same crop. One drawback to this procedure is that there were a large number of fields that were not classified because the technique needs p+1 data points in order to form the statistics required to assign it to a crop (where p is the number of bands or channels). However, if enough points are present, classification is possible.

Table 3--Per field classification matrix based on data from 3 overflights1/

Group	No. of fields	Percent fields correct	No. of pixels	Percent pixels correct	Cotton	Corn	Soybeans	Grass	Winter wheat	Odd	Not classified
Cotton	38	63.2	927	85.0	24	0	2	0	1	0	11
Corn	7	14.3	558	20.7	0	1	0	1	1	0	4
Soybeans	58	25.9	852	44.2	9	3	15	3	7	1	20
Grass	31	9.7	240	29.6	3	1	1	3	2	0	21
Winter wheat	5	40.0	85	56.5	1	0	0	1	2	0	1
Odd	4	50.0	55	80.0	0	0	1	0	0	2	1
Totals	143	32.9	2217	60.4	37	5	19	8	13	13	58

1/ August 26, 1972, MSS bands 4,5,7.
September 14, 1972, MSS bands 5,7.
October 2, 1972, MSS bands 4,5,6,7.

In the work we have done in Missouri, the sample classifier, about 40% of the fields were not classified because the required number of pixels for the classifier exceeded the number of pixels present within the defined fields. For the technique employed, 10 pixels per field were required.

In Missouri, 71% of the fields were less than 20 acres, but account for 32% of the total area. In our Kansas site, 20% of the fields were less than 20 acres, but account for only 1.5% of the total land area. In South Dakota, 40% of the fields were less than 20 acres, and accounted for 15% of the area. In idaho, 74% of the fields were less than 20 acres, and account for 25% of the area. If 20 acres is a critical field size for the classifier, we would expect to do well in making acreage estimates in Kansas, but in Missouri only a little more than 50% of the acreage would be accounted for.

The next table is a classification done on a single ERTS flight. For each ERTS pass there are 4 bands or channels of information. Three dates were overlaid, however, 3 out of 12 channels were of very poor quality and were unusable. Of the total 9 usable bands, 3 came from an August 26, 1972 pass, two from the September 14, 1972 pass, and 4 from the October 2, 1972 pass. Each point on the ground then has 9 different readings. Now to evaluate the information gained from the temporal overlay, compare Table 1 with Table 4. The gain is substantial using the information from the three passes.

Both comparisons indicate the gain for temporal information is about 10%. Also, if we compare Table 4 with Table 5, we find that the gain for using unequal prior probabilities over equal prior probabilities is 10%.

The results that we have presented up to now have been biased because we have used the same data for both training and testing. This procedure produces a classification table that shows better results than one should expect from independent or uncorrelated data. Figure 1 shows what the data looks like in two-space.

Table 4--Classification matrix for September 14, 1972 based on MSS bands 5 and 7.

Group	No. of sample points	Percent correct	Number of samples classified into					
			Cotton	Corn	Soybeans	Grass	Winter wheat	Odd
Cotton	927	71.4	662	44	36	47	116	22
Corn	58	34.5	12	20	6	9	2	9
Soybeans	852	28.9	184	62	246	132	210	18
Grass	240	44.6	43	21	45	107	22	2
Winter wheat	85	68.2	6	12	0	9	58	0
Odd	55	47.3	3	16	0	2	8	26
Totals	2217		910	175	333	306	416	77

Overall performance 50.5

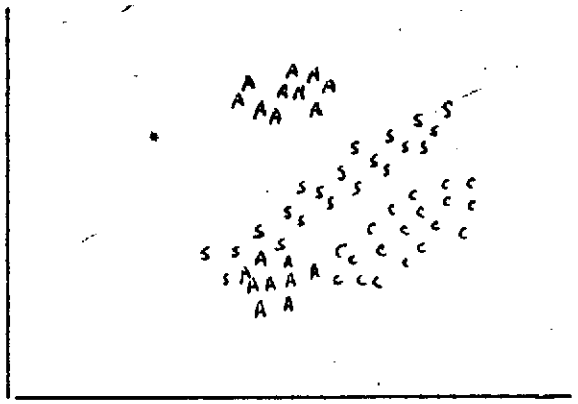
Another example is to compare Table 5 with Table 2.

Table 5--Classification matrix using September 14, 1972 MSS bands 5 and 7 with unequal prior probabilities.

Group	No. of sample points	Percent correct	Number of samples classified into					
			Cotton	Corn	Soybeans	Grass	Winter wheat	Odd
Cotton	927	69.7	646	0	246	14	0	21
Corn	58	0.0	12	0	16	20	0	10
Soybeans	852	67.6	175	1	576	74	0	26
Grass	240	42.1	40	0	97	101	0	2
Winter wheat	85	0.0	8	0	69	6	0	2
Odd	55	54.5	6	2	13	4	0	30
Totals	2217		887	3	1017	219	0	91

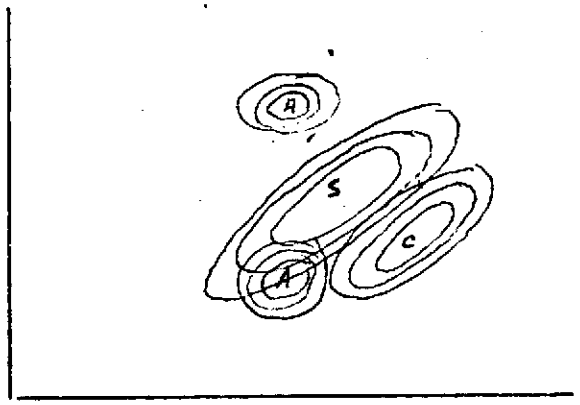
Overall performance 61.0

Figure 1--A graph of data obtained from two bands for Corn (C), Soybeans(S), and alfalfa (A).



A quadratic discriminant procedure assumes that the data was sampled from a multivariate normal distribution. Therefore, the sufficient statistics are formed, that is the mean vector and the covariance matrix. All other information about the population parameters is independent of these two statistics.

Figure 2--



To evaluate the discriminant functions, we need independent data if we are to get unbiased estimates. If the data set is large enough, the estimates will be very close to the population parameter. This is the property of consistency. When the estimates are close to the population parameters, the classification table of independent data coverages to the classification of non-independent data.

In order to measure the bias from using the same data as both training and testing, we tried a jacknife procedure. We divided the table into thirds. Two-thirds of the data was used as training data and the other one-third was used as test data. The results of this procedure are shown in Table 6. Naturally, we were not satisfied with the 34% correct classification, even though the results were free of bias. The results of the previous classification may be more accurate than this last procedure because 2/3 of the data sets may not be enough data to estimate the parameters.

Table 6--Classification matrix using August 26, 1972 data MSS bands 4,5, and 7 with independent test data.

Group	No. of sample points	Percent correct	Number of samples classified into					
			Cotton	Corn	Soybeans	Grass	Winter wheat	Odd
Cotton	900	55.3	498	69	141	77	72	43
Corn	58	27.6	4	16	2	13	9	14
Soybeans	852	15.0	89	88	111	174	367	23
Grass	240	28.3	36	45	19	92	45	3
Winter wheat	85	11.8	5	5	38	21	10	6
Odd	55	67.3	1	3	9	2	3	37
Totals	2190		633	226	320	379	506	126

Overall performance 34.6

The comparable classification where non-independent data was used is shown in Table 7.

Group	No. of sample points	Percent correct	Number of samples classified into					
			Cotton	Corn	Soybeans	Grass	Winter wheat	Odd
Cotton	927	60.7	563	92	108	63	58	43
Corn	58	56.9	2	33	0	7	11	5
Soybeans	852	15.3	57	72	130	245	322	26
Grass	240	45.4	32	41	26	109	29	3
Winter wheat	85	51.8	5	6	10	15	44	5
Odd	55	69.1	6	4	3	3	1	38
Totals	2217		665	248	277	442	465	120

Overall performance 41.4

Anytime the results differ this much between data sets, we know the data set is either too small or the bias is large. Obviously, we have not reached the point where we have covergence of parameters based on independent and non-independent data sets. The point is that the sample size necessary depends on the variation in the data set and the variation in the data set is generally a function of how dispersed the data really is. One thing is certain with a small data set, either procedure may lead to erroneous conclusions.

Classification of data at the Idaho test site is nearly complete. The results are based on 42 segments in the intensive agriculture strata in one ERTS frame. Two additional segments are not on this frame. The frame that contains these two segments also contains ten segments which are on the first frame so we may be able to use this overlapping data to calibrate from one frame to the next, or to measure the difference, due to frames in the means and variance for the overlapped data. A method of using calibration or training data in one frame to adjust parameters or to classify on another frame would be valuable since it would increase the value of the segment data. A crop may be different over a large area because of variety, fertilized soil type, weather conditions and stage of maturity, rather than technical factors associated with acquiring imagery. This may be possible in some areas and this problem should be investigated.

The data has serious banding problems. The problems seem to be most apparent in band 6 so that band was left out in the first classification.

Table 8-- Preliminary Classification of Idaho study area data using August 1972 data bands 4, 5, and 7 and unequal prior probabilities.

			Number of samples classified into											
	No. of Samples	Percent Correct	PEAS BEANS	HARV BEANS	BRLY	ALFALFA	CORN	FALOTH	IDLE	OHAY	PASTURE	SUGBTS	POTATOES	SPWH
Peas and Beans	579	14.5	84	45	1	31	0	0	0	0	327	89	2	0
Harvested Beans	784	71.1	13	562	45	8	0	0	0	0	152	4	0	0
Barley	1019	11.5	33	271	117	27	0	2	6	0	489	64	10	0
Alfalfa	1318	17.3	57	51	2	228	0	0	6	0	527	422	25	0
Corn	542	0.0	10	21	9	119	0	0	0	0	221	161	1	0
Fallow and Other	684	0.4	14	13	3	14	0	3	33	0	575	26	3	0
Idle	206	26.7	4	10	0	1	0	1	55	0	135	0	0	0
Other Hay	11	9.1	0	0	0	0	0	0	0	0	5	3	2	0
Pasture	1484	80.7	38	25	4	78	0	2	49	1	1197	83	8	0
Sugar Beets	527	76.5	12	6	1	43	0	0	6	0	46	403	10	0
Potatoes	533	10.1	29	2	1	80	0	0	0	0	89	278	54	0
Spring Wheat	111	0.0	3	48	3	5	0	0	0	0	49	3	0	0
Total	7798		297	1054	186	634	0	8	155	1	3812	1536	115	0

Overall Performance 34.7

Obviously, the classification is not as good as we expected, however, by chance one would expect only 8% correct classification. Another possible problem with the classification is that some field boundaries are located adjacent to other fields.' This means that the boundaries sometimes fall on adjacent points and since the pixels are partially overlapping these border pixels may be causing some of the problem. We will be looking at this more closely. The gray scale printout which follows, illustrates this problem.

Figure 3--Gray scale printout of a segment showing how fields are defined.

The next classification matrix uses equal prior probabilities.

The overall classification performance goes to 21.8%. This points out that prior information in terms of probabilities is also important in this test area.

The next classification was done to try to improve the performance. Some of the fields were redefined so that bad bands could be used. Also, some fields were redefined to eliminate border problems. Table 10 shows these results.

The overall performance has improved to 40.3% the effort was somewhat successful, but results are still poor. It seems that many crops are not distinct. Pasture seems to be mixed with every other crop. This is because the variances of the measurements from the pasture crop are large.

Table 9 --Preliminary Classification of Idaho study area data using August 1972 data bands 4,5, and 7 with equal prior probabilities.

	No. of Samples	Percent Correct	PEAS BEANS	HARV BEANS	BRLY	ALFALFA	CORN	FALOTH	IDLE	OHAY	PASTURE	SUGBTS	POTATOES	SPWH
Peas and Beans	597	25.6	148	43	1	29	19	26	109	96	12	25	59	12
Harvested Beans	784	66.1	20	518	40	15	4	18	50	7	8	1	14	89
Barley	1019	9.9	62	214	101	13	19	66	112	59	71	14	78	210
Alfalfa	1318	10.7	119	47	11	141	51	26	80	172	108	115	428	20
Corn	542	1.7	28	18	11	62	9	41	36	56	17	41	198	25
Fallow and Other	684	12.1	23	7	6	5	7	83	416	23	33	5	35	41
Idle	206	70.4	9	4	0	1	1	24	145	3	4	0	0	15
Other Hay	11	72.7	1	0	0	0	2	0	0	8	0	0	0	0
Pasture	1484	8.0	105	15	17	70	14	117	606	54	119	36	148	183
Sugar Beets	527	19.9	3	3	2	18	8	0	8	142	4	105	226	8
Potatoes	533	56.8	10	2	2	25	6	1	4	105	2	72	303	1
Spring Wheat	111	19.8	3	38	0	10	4	6	4	8	5	1	5	22
Total	7798		536	909	191	309	144	408	1570	733	383	415	1494	626

Overall performance 21.8

Table 10--Classification matrix of Idaho study area, August 1972 imagery using MSS bands 4,5,6, and 7.

	No. of Samples	Percent Correct	PEAS BEANS	HARV BEANS	BRLY	ALFALFA	CORN	FALOTH	PASTURE	SUGBTS	POTATOES	SPWH
Peans and Beans	549	40.6	223	6	9	23	4	61	123	94	5	1
Harvested Beans	813	62.6	19	509	106	11	1	38	121	6	0	2
Barley	957	75.9	68	108	248	65	9	83	331	36	6	3
Alfalfa	1314	29.8	192	30	34	391	30	32	331	254	23	1
Corn	541	8.5	42	13	20	106	46	52	186	69	8	4
Fallow and Other	779	37.4	28	1	7	31	3	291	412	3	3	0
Pasture	1433	64.0	107	8	24	115	8	218	917	34	2	0
Sugar Beets	386	56.0	19	1	5	60	8	1	30	216	45	1
Potatoes	395	21.8	15	0	0	115	7	0	92	80	86	0
Spring Wheat	104	3.8	12	27	24	4	1	3	23	4	2	4
Total	7271		725	703	477	921	117	779	2566	787	180	16

Overall performance 40.3

APPENDIX A

PDSCMS

A PROGRAM TO CONVERT PDS MICRODENSITOMETER SCAN LINES INTO SAS COMPATABLE OBSERVATIONS

by

Paul D. Hopkins

January 7, 1974

This program is designed to convert a PDS microdensitometer scan into a SAS compatable multivariate observation. Up to 4 scans of the same area may be included in the SAS observation.

The user controls the number of scans (normally 1 for each filter) to be used in building the multivariate observations. The microdensitometer scans are read in serially and saved on temporary files. After all the data for a given pisect has been read in, the temporary files are rewound and read back a line at a time, and a SAS observation produced for each point in the line. Each observation consists of data from corresponding points from all scans used.

The program is divided into 3 phases: (1) parameter phase, (2) read phase, and (3) combine phase. The normal operation of the program is to go from phase 1, to phase 2, to phase 3, and repeat as desired.

Parameter phase:

Allows the user to define the initial settings for all counters, and indicators used during the read and combine phases. If fatal errors occur during the run, control reverts to the parameter phase for an error scan of all remaining control cards, but no data will be processed.

Read phase:

During the read phase microdensitometer scans are read in and stored on temporary files. During this process, the PDS 9-tract format is converted to a 8 bit internal IBM notation. If the data was scanned in a raster or right edge scan, it is converted to a left edge scan. The user, however, may elect to cancel this option and accept the data in the order scanned. While in read phase, all parameter definition cards are ignored. If an attempt is made to read more than 4 scans, the combine phase is automatically entered.

Combine phase:

This phase combines the results of the read phase. Corresponding points from each read file are included in each SAS observation produced. The data from the reads are put in correspondence with the data items in the SAS observation set. If these are fewer than 4 scans to be combined, the trailing data items are assigned the missing value. The cordinate values and pixel serial numbers are computed and assigned as each observation is produced. At the conclusion of this phase, control reverts to the parameter phase, and new parameter settings will be accepted.

NUMERIC VALUE REPRESENTATION

The microdensitometer output is a ditigal representation of an analog signal. The amount of light pasing through a sample is converted into a voltage by a photo-multiplier tube. If transmissions are being recorded, the voltage is routed to the panel display meter and then to the A/D (analog to digital) converter. If optical densities are being recorded, the voltage is first sent to a logarithmic converter before goint to the panel display meter and then to the A/D converter.

The A/D converter produces a positive integer value that represents the voltage. The input range of the A/D converter is 0.00 to 5.12 volts in .005 volt increments. The digital output ranges from 0 to 1024, or 200 times the voltage input. It is important to remember that these values could be either transmission or density depending on the calibration settings.

When the digital output from the A/D converter is stored in the computer (PDP8), it is multiplied by 2 and is now 400 times the value shown on the panel meter. This is done to reduce the effect of noise contamination. Some noise could result from the fact that the microdensitometer actually takes discrete readings from a continuously varying function.

The data values are recorded in a 9-track tape format. The PDP8 computer is a 12 bit word machine with 6 bit bytes and is not directly compatable with the 9-track 8 bit byte tape format. Therefore, 2 zero pad bits are appended to each PDP8 byte as it is written in a 9-track format. Physically, the data on tape has the format shown below:

ppsdddddppdddddn

where p represents the pad bits appended to fill the 9-track tape format,
s is the PDP8 sign bit and is normally 0,
d represents one of the 10 data bits from the A/D converter,
n represents the noise bit position, normally 0.

In reconstructing the microdensitometer data back into a useable form, the program allows the user two choices. By default, values will be produced from storage type data. Optionally, actual panel display values may be generated.

Storage data has been reduced to a form which is suitable for bulk storage. Each value is reduced to an 8 bit integer and requires exactly 1 byte of storage. This is the form used by ERTS, LARSYS, and the Penn State Classification System.

The numeric range of the integer volved data is from 0 to 255. Approximate panel values may be derived by multiplying a storage value by .02. At first, it may seem that we are discarding valid data, but this is not so if we consider the accuracy of the microdensitometer.

The microdensitometer specifies linearity of ±.02 density or .5% transmission, and that the drift for a 10 hour period is less than ±.02 density or less than 1% transmission. This means that a recorded value could differ from the true value by as much as .04 density or 1.5% transmission. The stored values will resolve density to the nearest .02 units and transmission to the nearest .4% (.3921569), which is within the limits of the equipment.

The Panel Data option allows the reconstruction of exact panel readings as shown by the panel display meter. The data accuracy implied is beyond the capability of equipment, but it should be useful in checking machine specifications.

X Y COORDINATE SYSTEM

The program assumes a generalized coordinate reference system. The x,y coordinates are signed integers, with (0,0) as the default origin. The x ordinate is the element index, and the y ordinate is the line index. The program always assigns thealgebraically smallest x,y value to the pixel in the north-west corner (upper left). The x ordinate increases as the scan moves to the east (right), and the y ordinate increases as the lines move south((down).

The PDS microdensitometer normally scans lines in a raster (back & forth) with the direction of scan alternating, and can scan lines from top to bottom or bottom to top. The PDSCMS program has the ability to determine the scanning directions, and use this in the coordinate assignment algorithm. Thus, regardless of how the points are scanned, the above defined coordinate reference system is valid.

The program computes the coordinates during the combine phase. The coordinates of the physically first point are computed and assigned to that point. If this point is not the north-west corner point, the coordinate of the north-west corner point are derived. The program prints out the north-west corner coordinates as the first x and y ordinates.

The above described coordinate reference system may seem unduly complicated, but it (1) sets up a reference system that is both hardware and software compatable, and (2) permits full use of the microdensitometer scanning ability.

Display devices such as line printers and CRT devices, display data from left to right and top to bottom. The natural order of computer indexing is from smallest to highest. Thus, after coordinates are assigned, data points may be sorted by coordinate and they will be in the natural order for computer processing regardless of how scanned.

The user may have several scans from a scene with the microdensitometer defining the origin at each pisect. The conversion software would call that point (0,0) by default. Later, the user may wish to restore or assign relative position of pisects by relocation. The user could also move the origins of all pisects from the microdensitometer (0,0) setting to any arbitrary point (n,n).

The user may have the microdensitometer scan several pisects from a scene relative to a common origin. The conversion softwarewill compute initial coordinates for each pisect using the microdensitometer supplied locations. Thus, the resulting pixel coordinate will preserve the relative saptial location of the pisects relative to thescene origin. Later, the user may wish to perform an origin transformation, and spatially relocate this scene relative to any other independently scanned scene.

SAS OBSERVATIONS

Each observation produced has 11 items as follows:

SCENE-NAME — 1-8 characters left justified with trailing blanks in bytes 5-12.

This name is used to identify a collection of pisects (picture sections). If the user fails to supply a valid name, the program will use the current date in the form mm/dd/yy by default.

PISECT-NAME — 1-8 characters left justified with trailing blanks in bytes 13-20.

This name is used to identify a pisect within a scene. A new name is supplied for each pisect processed. If the user fails to supply a valid name, the program will use the current value of the system clock in the form hh•mm•ss by default.

GROUP-NAME — 1-8 characters left justified with trailing blanks in bytes 21-28.

This name is used to identify calibration data. A null or 'blank' name indicates unknown data. The discriminate function, uses named groups as training, and classifies unknown data. If the user fails to supply a valid name, the program supplies the null or 'blank' name by default.

IDENT-NAME — 1-8 characters left justified with trailing blanks in bytes 29-36.

This name is used to establish user identity of unknown data. A null or 'blank' name indicates that the user does not know or cannot identify the item. Valid ident-names are taken from the set of group names. The discriminate function would use the ident-name to check classification accuracy. If the user fails to provide a valid name, the program supplies the null or 'blank' name by default.

XORD — integer binary in bytes 37-40.

This is the relative position of the SAS observation within a line of data. It always gives relative element position within its own pisect, and depending on user options may be positional relative to an entire scene or group of scenes.

YORD — integer binary in bytes 41-44.

This is the relative line position of the SAS observation. It always gives relative line position within its own pisect, and depending on user options may be positional relative to an entire scene or group of scenes.

PSN integer binary in bytes 45-48.

This is the pixel serial number assigned by the program. Pixels are serialized in order processed in the combine phase. Unless directed otherwise, pixels are serialized for the entire run starting with 1. The serial number may be signed.

PIXF1V real binary in bytes 49-52.

This is the microdensitometer value for the first scan read for the current pisect. It will never be assigned the missing value.

PIXF2V real binary in bytes 53-56.

This is the microdensitometer value for the second scan read in for the current pisect. If there was no second scan, it takes on the missing value.

PIXF3V real binary in bytes 57-60.

This is the microdensitometer value for the third scan read in for current pisect. If there was no third scan, it takes on the missing value.

PIXF45 Real binary in bytes 61-64.

This is the microdensitometer value for the fourth scan read in for the current pisect. If there was no fourth scan, it takes on the missing value.

The program writes the SAS compatable file in binary (unformatted) variable blocked spanned mode. (RECFM=VBS). Because SAS includes the record description word as part of the record, the byte locations of all items have been offset by 4 bytes in the above description.

CONTROL CARDS

The program uses 11 different control cards. Most of them are optional because the program will supply default values when the user does not. Each control card is divided into 3 major fields as follows: (1) key word or op-code in columns 1-8; (2) parameter field in columns 11-50; and, (3) comments field in columns 51-80.

There are 3 classes of control cards, depending on the kind of action to be performed. Each class is described separately below:

Class 1 - Run Cards

These cards set indicators that remain in effect for the duration of the run or until redefined during the run. All run cards are optional.

SCENE Card

cols 1-8 SCENE

cols 11-18 1-8 character name left justified with trailing blanks used to identify a group of pisects. The contents of columns 11-18 are placed in the scene-name field of the SAS compatable record. If the user does not make a scene, the program supplies the current date by default.

PSN Card

cols 1-8 PSN

cols 11-15 signed integer constant starting serial number.

This card can be used to extend the serialization of previous computer runs. If the user does not supply a starting serial number, a value of 1 will be used by default.

ORIGIN Card

cols 1-8 ORIGIN

cols 11-15 signed integer constant x cordinate offset.

cols 16-20 signed integer constant y cordinate offset.

This control card is used to provide origin translation of each pisect processed. The cordinates of the first point are computed and the offset applied. It may be used to relate the pisects from the current scene to those in a previous or subsequent scene. This feature may be useful when the data are from sequential scenes such as aircraft photography.

If the user does not supply an origin card, the 0,0 or no transformation will be done.

EDGE Card

cols 1-8 EDGE

This card causes the program to convert to all scans to a left edge scan. This effectively removes the raster produced by the back and forth microdensitometer scanning motion. All lines running from right to left are turned around. If an EDGE card is not supplied, it is assumed.

ASIS Card

cols 1-8 ASIS

This card causes the program to accept the data points in the order scanned. However, the x,y cordinate assigned are computed based on line direction. If the pixels are sorted based on the x,y coordinates,a normal picture will be produced. That is, the true northwest corner point has the algebraically smallest coordinates,and the southeast corner has the algebraically largest coordinates. If an ASIS card is not supplied, EDGE is assumed by default.

ABL Card

cols 1-8 ABL

This card causes the program to accept microdensitometer data sets that have identified with blank or first character blank labels. By default such scans are rejected as a fatal error. Note that once turned on this option cannot be rescinded during a computer run.

VALUE Card

cols 1-8 VALUE

cols 11-18 STORAGE
PANEL

This card allows the user to select the type of numeric values to produce for the SAS file. Storage values are normalized floating point integers, range $0 \leq$ value ≤ 255. Panel values are also normalized floating point, but is the microdensitometer A/D converter output expressed as a display panel number. The range is $0.000 \leq$ values ≤ 5.115, in increments of .005. A storage value is numerically 50 times the panel value with the decimal fraction truncated.

When a value card is used, and the identifier in columns 11-18 are not PANEL, storage values are produced by default.

Class 2 - Pisect Cards

These cards set parameters that apply only to the pisect about to be processed. They are automatically cleared to default values after a COMBINE control card. All pisect cards are optional.

PISECT Card

cols 1-8 PISECT

cols 11-18 1-8 character name left justified with trailing blanks.

The contents of columns 11-18 are saved in the pisect-name in the SAS Compatable record. It serves to identify pisects within scenes. If the user does not supply a PISECT card, the program uses the current value of the system clock by default.

GROUP Card

cols 1-8 GROUP

cols 11-18 1-8 character name left justified with trailing blanks.

The contents of columns 11-18 are placed in the group field in the SAS Compatable record. A non-blank name indicates that this pisect contains calibration data for a specific group. If the user does not supply a group name, the program inserts a blank name by default.

IDENT Card

cols 1-8 IDENT

cols 11-18 1-8 character name left justified with trailing blanks.

The contents of columns 11-18 are placed in the ident-name field of the SAS Compatable record. A non-blank name indicates that the user has identified the points in this pisect as belonging to the specified group. If the user does not supply an IDENT, the program inserts blanks by default.

RELOCATE Card

cols 1-8 RELOCATE

cols 11-15 signed integer constant representing the north-west x ordinate.

cols 16-20 signed integer constant representing the north-west y ordinate.

The north-west corner pixel will be assigned the cordinates given on this card. All subsequent pixels will be assigned cordinates relative to these. Thus, any pisect can be arbitrarilly moved in space. By default, absolute relocation will not be performed.

This card overides the origin transformation in effect for each pisect for which relocation is performed. The origin transformation will be performed for each pisect not relocated.

Class 3 - File Manipulation Cards

These control cards cause data to be moved from one file to another, and to perform some transformations on the process. These cards are required as specified below.

READ Card

cols 1-8 READ

cols 11-50 1-40 character names left justified with trailing blanks.

This card causes the program to read in 1 PDS microdensitometer scan to be read in, stored on a temporary file. One read card is required for each scan to be included in a SAS observation. When a read card is processed, while the program is in the parameter phase, control is switched to the read phase. No more parameter cards will be honored until control reverts back to the parameter phase.

Up to 4 consecutive read cards will be honored. If a 5th read card is encountered, the program will combine the 4 scans already stored on temporary files, and then scan the remaining control cards for errors. No more data will be transferred. Either an end-of-file or a combine card must follow read cards.

The 1-40 character name is used for label checking as follows:

(1) If the name is absent or begins with a blank the program assumes that no label checking is to be performed, and whatever file it finds is assumed to be correct.

(2) If a name is present, it must match the label put in the scan line by the microdensitometer operator. Label checking is performed up to the first blank character in the supplied name. Thus, if the user has no common prefix for a series of scans, he may use an abbreviated label to verify that the correct scans are being processed. If the label check fails, no more files are processed, but the remaining control cards are checked for errors.

COMBINE Card

cols 1-8 COMBINE

This card causes the program to combine the results of the previous reads and add the results to the SAS compatable data set being built. If n scans are being combined, exactly n-1 combine cards are required. The last combine card in the control card stream is optional as any uncombined reads are automatically combined at end-of-file. At the end of a combine operation, the program returns to the parameter phase and will accept parameter control cards.

EXECUTING THE PDSCMS PROGRAM

The PDSCMS Program is executed by using the RADLGO procedure. The PDS microdensitometer tape is read in from unit 8, and the converted file is written on unit 9. Program control cards are read from SYSIN.

The microdensitometer output is a series of stacked data sets on magnetic tapes. The program reads are many data sets from the stack as directed by READ control cards and incrementing the unit 8 FORTRAN Sequence Number. Each READ control card requires a unit 8 DD JCL statement with an appropriate sequence number. The data set sequence number in the label parameter points to the particular scan to be processed by the READ command.

```
//FT08F001  DD  LABEL=(i,NL,,IN)        for first READ card
//FT08F002  DD  LABEL=(j,NL,,IN)        for second READ card
//FT08F003  DD  LABEL=(k,NL,,IN)        for third READ card
     .
     .
     .
//FT08Fnnn  DD  LABEL= (m,NL,,IN)       for nnn'th READ card
```

The letter i,j,k,m, represents the data set sequence number on the tape and point to the i'th, j'th, k'th, and m'th data set respectively.

The converted SAS file is written on unit 9 in FORTRAN binary (unformatted) mode as a single unstacked data set.

SAMPLE JCL

```
//XO  EXEC  RADLGO,                          load & execute
//          P=PDSCMS                         the PDSCMS program
//GO.FT08F001  DD  DISP=OLD,UNIT=2400,DCB=BLKSIZE=6400,RECFM=U,BUFNO=1),
//                 VOL=SER=URxxxx,
//                 LAVEL=(i,NL,,IN)
//GO.FT08F002  DD  DISP=OLD,UNIT=2400,DCB=*.FT08F001,VOL=REF=*.FT08F001,
//                 LABEL=(j,NL,,IN)
//GO.FT08F003  DD  DISP=OLD,UNIT=2400,DCB=*.FT08F001,VOL=REF=*.FT08F001,
//                 LABEL=(k,NL,,IN)
     .                                                          .
     .  as many dd statements as needed; extra ones do no harm.
     .                                                          .
//GO.FT08Fnnn  DD  DISP=OLD,UNIT=2400,DCB=*,FT08F001,VOL=REF=*,FT08F001
//                 LABEL=(m,NL,,IN)
//GO.FT08F001  DD  DSN=dsname,DISP=(,KEEP),UNIT=2400
                   DCB=(BLKSIZE=6400,LRECL=32000,RECFM=VBS,BUFNO=1)
//GO.SYSIN     DD  *
     PDSCMS control cards
/*   EOJ.
```

SAS PROCESSING THE COMPATABLE FILE

JCL Requirements

In order to process the compatable file with the SAS program, an additional DD statement is required by the RADSAS procedure. This statement is required to point to the file to be used. In the following JCL, the PDSFILE[1/] DD statement is used to gain access to the converted PDS data.

```
//S  EXEC  RADSAS
//PDSFILE  DD  DSN=dsname,DISP=OLD,UNIT=2400,VOL=SER=xxxxxx
//SYSIN    DD  *

      .
      . sas program statements
      .

/*   EOJ.
```

In the above example, the converted file is assumed to reside on magnetic tape. If the file is not on magnetic tape, or is passed from a previous job step, an appropriate alternation in the PDSFILE DD statement will be required.

SAS Program Statements

The SAS program must be directed to use the PDSFILE DD statement for its input. The model statements given below can be used to read in all the items from the converted file.

```
DATA:
INPUT  DDNAME=PDSFILE  SCENE $ 5-12  PISECT $ 13-20  GROUP $ 21-28
IDENT $ 29-36  XORD  IB 37-40  YORD  IB 41-44  PSN  IB 45-48
PIXF1V  RB 49-52  PIXF2V  RB 53-56  PIXF3V  RB 57-60  PIXF4V  RB 61-64
```

The user may not wish to read in all the items. Those items not wanted may be omitted from the list in the input statement. The following statement shows how to read in only the data from the first and third read cards.

```
DATA;
INPUT  DDNAME=PDSFILE  PIXF1V  RB 49-52  PIXF3V  RV 57-60;
```

1/ The user may substitute any name for PDSFILE, but that name must also be used in the SAS INPUT statement.

The PDSCMS program assigns the missing value to the PIXFiV elements for which there was no corresponding read card. The user can do 1 of 4 things with missing value: (1) accept data with missing values and let SAS handle them, (2) do not read in the pixel filter values that are missing, (3) convert the missing value to some neutral value, or (4) identify and take special action for missing items.

Sample Program To Convert Missing Values to 0.

```
PIXF2V=PIXF2V+0;
PIXF3V=PIXF3V+0;
PIXF4V=PIXF4V+0;
```

Sample Program to Drop Missing Values.

The program examines the first record for missing values to determine how many items to drop. Thereafter, the same number of items are dropped from every record. Note also, that instead of dropping these items, any special values could be assigned, or special processing could be performed.

```
TDI:  IF DI < 0  THEN GO TO DDI;
      IF DI = 0  THEN GO TO SDI;
      IF DI = 3  THEN GO TO D234;
      IF DI = 2  THEN GO TO D34;
      GO TO D4;

SDI:  DI=-1;
      IF NO PIXF4V  THEN DI=1;
      IF NO PIXF3V  THEN DI=2;
      IF NO PIXF2V  THEN DI=3;
      GO TO TDI;

D234: DROP PIXF2V;
D34:  DROP PIXF3V;
D4:   DROP PIXF4V;
DDI:  DROP DI;
```

DATA CONVERSION

Microdensitometer data is expected to be used from a storage format which is an 8 bit integer value from 0 to 255 inclusive. Storage data can either represent densities (logarithmic response), or transmission (linear response). Simple linear transformations are required to reduce storage values into the corresponding panel meter value, optical density, or percent transmission.

Storage values can be converted directly into corresponding panel meter values by multiplying by .02.[1/] The resultant is either an optical density or transmission value, depending on the microdensitometer calibration settings when the scan was performed.

When the microdensitometer is calibrated to record densities, the panel value is optical density. Storage values are increments of .02 density units with a valid range from 0.00 to 4.00 inclusive. Density readings larger than 4.00 constitute an overflow condition because they are beyond the specified range of the equipment.

When the microdensitometer is recording transmissions, the stored data represents an incremental percent transmission that is dependent on the gain setting during calibration. Normally, the gain is set at 5.10 to give maximum range and accuracy to the transmission levels. The incremental step is then .3921569% transmission.

In addition, it may be useful to convert the storage data into, from logarithmic densities into linear transmissions and vice versa. In the following relationships, the transmission calibration (Gain) is assumed to be 5.10. The density is always calibrated to 0.

The following symbols are used in the equations that follow.

DS	density (logarithmic) storage value	$0 \leq DS \leq 200$
ST	Transmission (linear) storage value	$0 \leq ST \leq 255$
G	Gain setting for transmission	nominal value 5.10
PT	Percent transmission	$0 \leq PT \leq 100$
OD	Optical density	$0 \leq OD \leq 4.00$

1/ Described in the numeric representation section.

The relationship between optical density and transmission is:

$$\text{Density} = -\text{Log}_n(1/\text{Transmission})$$

If we impose on this basic relationship, the requirement that 100% transmission is 0 density and 0% transmission is 4.00 density, the equation can be rewritten as:

$$OD = 2 - \log_{10}(PT)$$

or

$$PT = 10 **(2 - OD)$$

Note that the relationship of 0% transmission = 4.00 optical density requires a mathematical impossibility, namely $\text{Log}_{10}(0) = -2$, and $10^{-2} = 0$. These conditions are definitional and are imposed by the resolution limits of the electronic circuting in the microdensitometer. During computer processing this limiting point requires special handling. Computationally, the valid conversion ranges for percent transmission and optical density are:

$$0 < PT \leq 100$$
$$4.00 > OD \geq 0$$

Also, be aware that 4.00 optical density can be transformed into the computionally valid percent transmission value .01. It storage transmissions are being produced the minimum storage value is .39% and is larger than .01. An attempt to produce a storage value for .0% transmission will result in a 0 value.

Because in the density to transmission, computations can be performed over the entire density range, it is possible to computationally extend the valid transmission range beyond 2.3 optical density. An image is digitized in densities and the corresponding percent transmission computed. Thus, a percent transmission values less than .39, can be used in computations, but cannot be produced by the microdensitometer, nor stored in standard form.

The equation to convert stored density data into optical density is:

$$OD = SD * .02$$

The equation to convert stored transmission data into percent transmission is:

$$PT = ST * .3921569 \qquad \text{when } G = 5.10$$
$$PT = ST * (2/G) \qquad 0 < G < 5.10$$

The following transformations are used to convert logarithmic values into linear values and vice versa.

To convert stored density into percent transmission use:

$$PT = 10 **(2 - SD* .02)$$

To convert stored density into stored transmission use:

$$ST = 10 **(2.40654 - (SD *.02)) \qquad G = 5.10 \text{ implied}$$

To convert Optical Density into stored transmission use:

$$ST = 10 **(2.40654 - OD) \qquad G = 5.10 \text{ implied}$$

To convert stored transmission into optical density use:

$$DS = (2 - \log_{10}(ST * .3921569)) * 50 \qquad G = 5.10$$

$$DS = (2 - \log_{10}(ST * (2/G))) * 50 \qquad 0 < G < 5.10$$

To convert percent transmission into stored density use:

$$DS = (2 - \log_{10}(PT)) * 50$$

APPENDIX B

LOCATING SEGMENTS ON ERTS IMAGERY

Locating Segments (Ground Truth) on ERTS Imagery

The segments used as ground truth or training data for this study were based on parcels of land chosen for enumeration by the Statistical Reporting Service. These segments of land are generally about one square mile in size. The use of this type of ground truth had a two-fold advantage. The first advantage is random selection. Segments chosen at random should be representative of the areas both as to crop and maturity. The importance of representative ground truth is vital when one is trying to classify very large areas. The second advantage is the ability to make estimates from the classification of the segments alone. The design of segment selection was made such that expansion factors and variance estimators for the sampling procedure are available.

In constructing the frame from which the area samples are drawn, the State is first stratified according to land use. After a state is stratified, each stratum is split into count units. A count unit is a specific area of land with an assigned number of sampling units. The sampling units are then chosen at random from the count units. The selected segments are drawn on county highway maps. We also have 24x24 and 9x9 ASCS aerial prints of the area where each segment is located. Segment boundaries are drawn in using a permanent marker. With 9x9 contacts in hand, field enumerators were sent to interview the farm operators in each segment. They were asked to draw in each field on the aerial print, give the field size and identify the crop. Visits are made during the crop year to check on the crops progress and to learn if fields were harvested and replanted to another crop. In this way, each field is identified for later processing.

During the growing season some color infrared aerial photography was taken of selected segments in the study. These flights were to be made on or about the same day as the satellite passes, but some were later.

In addition, we had a 38x38 blow-up of several ERTS scenes to have each segment included in at least one as ERTS Imagery print. This was done to have visible features to use when locating segments in from the MSS tapes.

The first task was to find the segment on the aerial photography. Landmarks found on the county highway maps such as lakes, major highways, airports and towns, etc., are used to find the general area. Then the fine details such as county roads and smaller streams are used to pinpoint the actual segment location. Generally, finding segments on the aerial photography was fairly straightforward.

Airplane flightlines are then drawn on the ERTS Imagery photo. The ERTS photo covers an area of about 10,000 square nautical miles and may contain one or more flightlines. Segments that are on the aerial film can be easily located on the ERTS imagery prints within the plane flightlines, providing that the plane and Satellite photos were taken within a couple of days of each other.

Segments that fall outside of the flightlines are more difficult to locate on the ERTS imagery photo. To locate these segments, measurements are taken between landmarks, or points which are visible on both, the ERTS imagery and county highway maps. The scale of each county map is located on the map somewhere, usually on the lower right hand corner. Some county maps are drawn like a grid (Squares). By counting squares or measuring, we can find the distance between segments or landmarks to the segment. A ratio between the two, county maps and ERTS imagery, is found then measurements between landmarks or segments on the county map are used to locate these points on the ERTS Imagery photo.

The ERTS Computerized Printout Sheet like any photography, has light and dark or shaded areas. These dark and light areas are printed on an 8 to 9 foot computer printout sheet with typing characters representing each pixel. Darkest areas are represented with #(no. or lb.) symbol. The next darkest areas are represented by @ symbol, followed by $,M,Y,/, the dash, period, and blank spaces represent the lightest areas on the ERTS imagery computerized printout sheet.

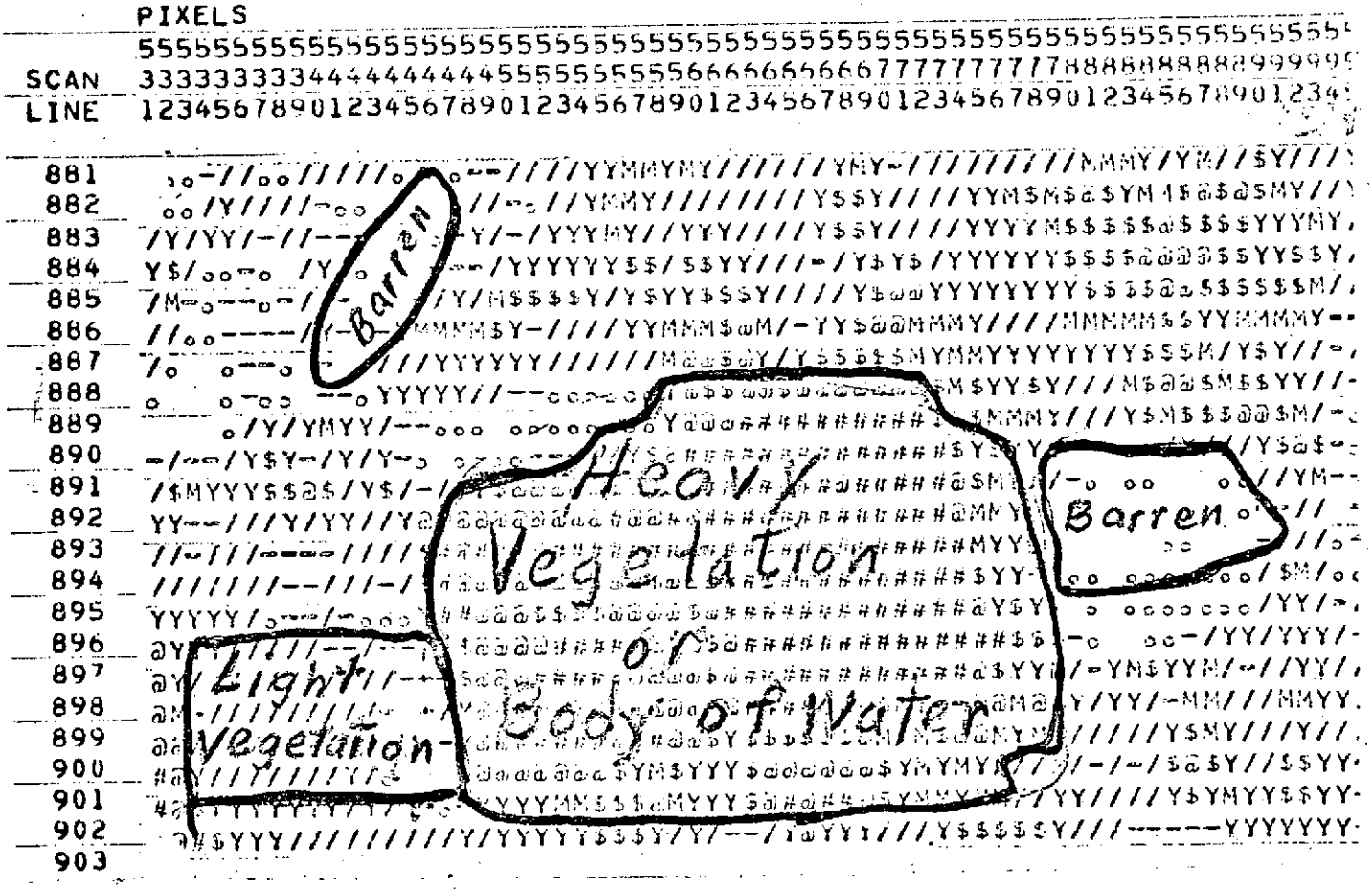
GRAY SCALE LEVELS:	
SYMBOL	
#	darkest area
@	next darkest area
$	" " "
M	" " "
Y	" " "
/	" " "
-	lightest area
°	next lightest area

A difficulty often encountered is that of obtaining sufficient contrast on the printout to determine the location of each field. Since the Penn State N-map program does not contain a histogram subroutine for setting the proper levels, often a second N-map had to be run after the class level cards had been readjusted to obtain a more even distribution of the grey-scale percentages into classes. There are 128 grey-scale levels which can be divided to the nearest .1% so that all 100% of the grey-scale levels on the printout are represented. The correct amount of contrast can usually be obtained by dividing an equal percentage of levels into each class used in the printout. For example, if 8 classes are used, then each class should be made up of 12 to 13% of the total. A program W-map was developed which will sample a given area and fix the grey-scale percentages into classes. This program helped to eliminate some duplicate mapping.

When the ERTS Computerized printout sheets are observed, the light and dark areas can be seen. The dark areas represent lakes, rivers, fields, or areas of heavy vegetation, forests, and cloud shadows (if clouds are present). The lightest areas represent barren lands, plowed farmlands, or areas of harvested crops, concrete highways, etc.

THE CHARACTER SET USED FOR DISPLAY IS

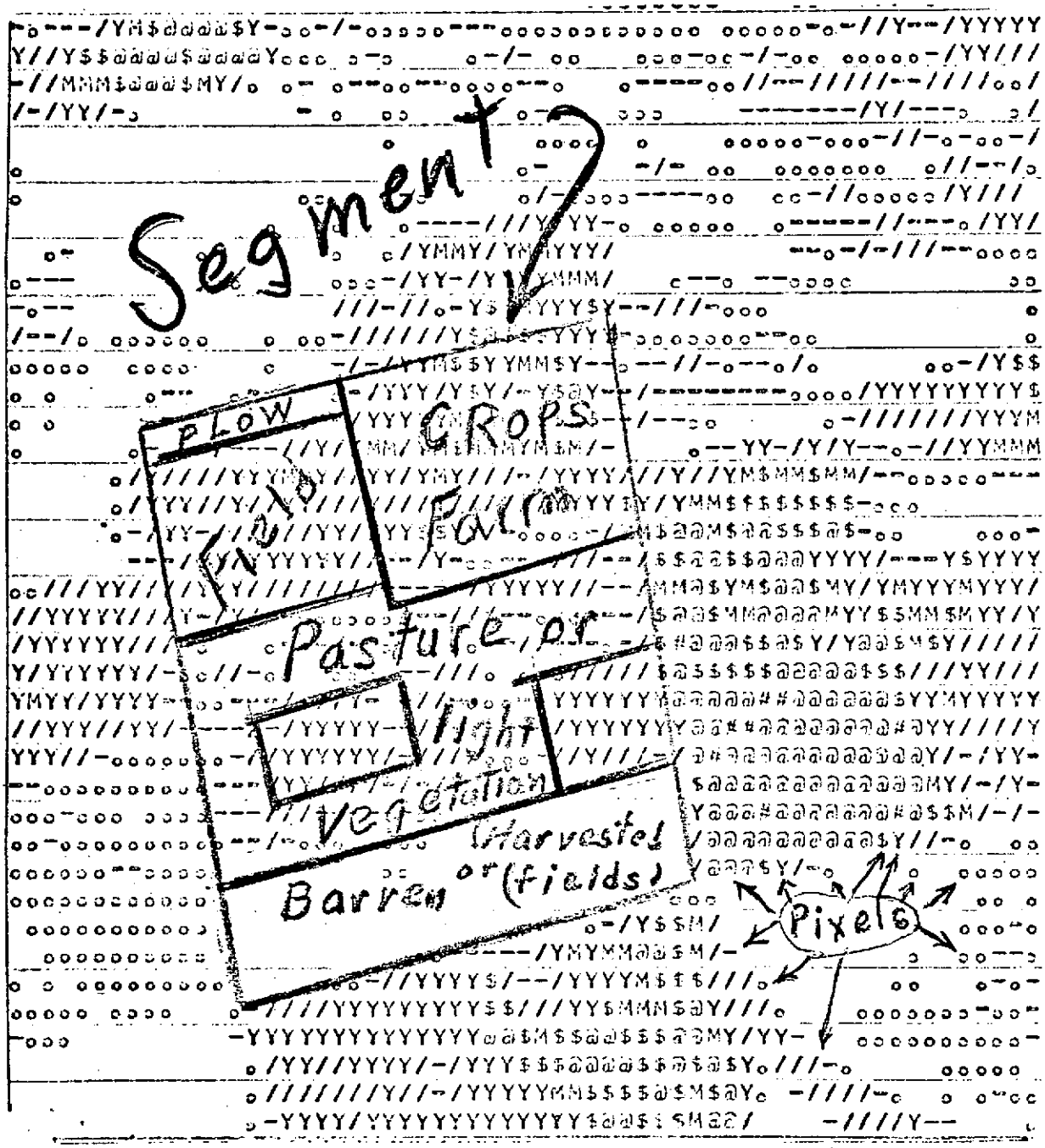
LEVEL	1	FROM	0.5	TO	13.5	DISPLAYED AS	#	4.900 PERCENT
LEVEL	2	FROM	13.5	TO	14.5	DISPLAYED AS	@	8.700 PERCENT
LEVEL	3	FROM	14.5	TO	15.5	DISPLAYED AS	$	13.967 PERCENT
LEVEL	4	FROM	15.5	TO	16.5	DISPLAYED AS	M	11.000 PERCENT
LEVEL	5	FROM	16.5	TO	17.5	DISPLAYED AS	Y	13.033 PERCENT
LEVEL	6	FROM	17.5	TO	18.5	DISPLAYED AS	/	5.400 PERCENT
LEVEL	7	FROM	18.5	TO	19.5	DISPLAYED AS	-	10.367 PERCENT
LEVEL	8	FROM	19.5	TO	20.5	DISPLAYED AS	.	12.533 PERCENT
LEVEL	9	FROM	20.5	TO	127.5	DISPLAYED AS		9.200 PERCENT



Dark areas such as lakes, rivers, circular irrigated fields, stand out as landmarks. By using a special template to measure the distance from these landmarks the location of the segments are found. With the aid of ASCA prints, aerial color IR photography, ERTS Imagery photo, ground truth all help to pinpoint the farm, field, or rangeland in the selected segment on the ERTS computerized printout (as shown following).

ERTS Computerized Printout Sheet Section

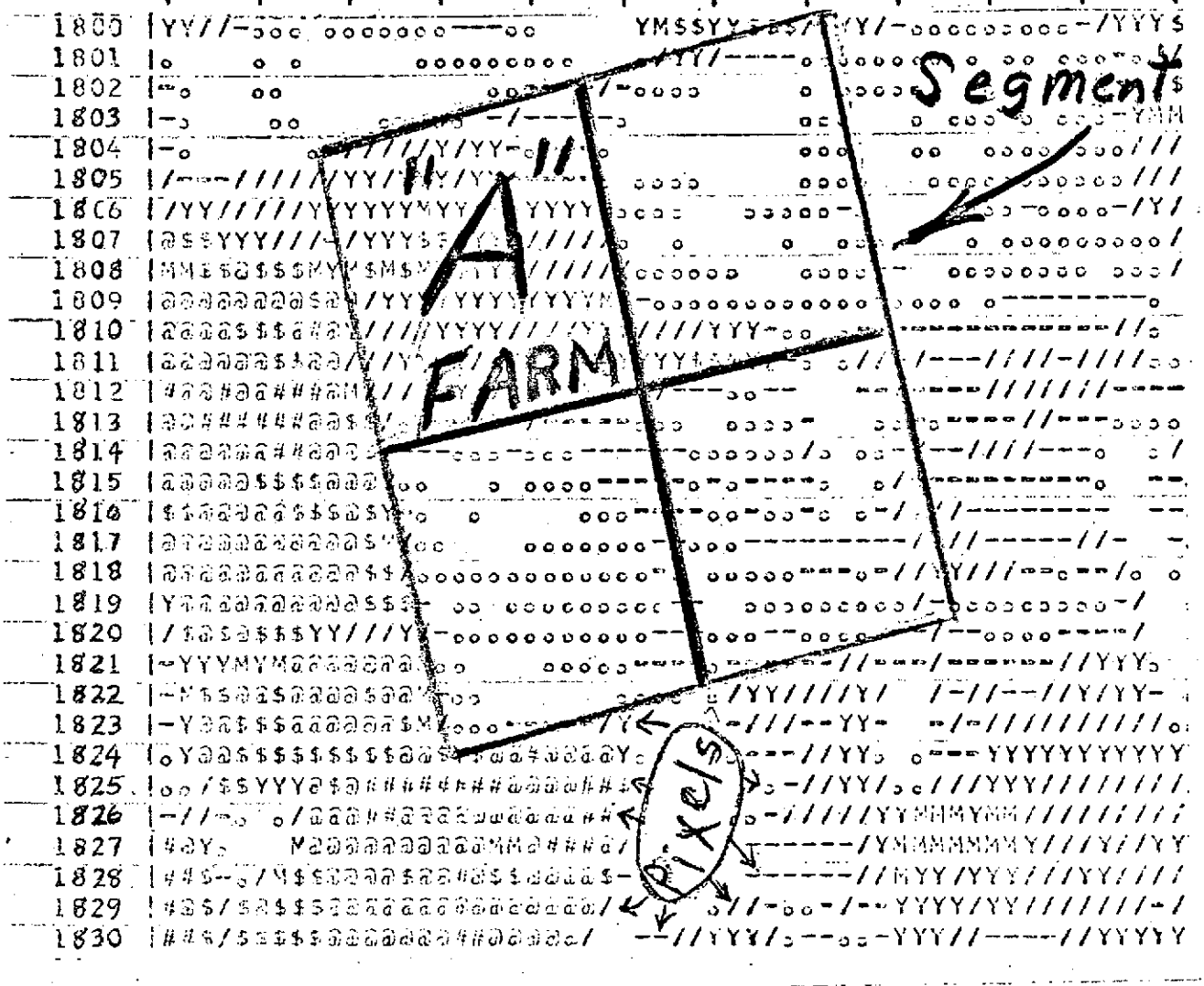
Including Segments



Imagery photo, ground truth all help to pinpoint the farm, field, or rangeland in the selected segment on the ERTS computerized printout (as shown above).

BLOCK SPECIFICATIONS

BEGINNING LINE	~~1675~~ 1800
ENDING LINE	~~2338~~ 1830
BEGINNING ELEMENT	2337
ENDING ELEMENT	~~2424~~ 2392
LINE INCREMENT	1
ELEMENT INCREMENT	1



The light and dark symbols are known as pixels. There are approximately 588 pixels to one square mile (640 acres) as represented on the computer printout sheets.

1 pixel = approximately 1.088 acres - Each acre represents about .919 pixels

1	acre	=	approximately	0.919	pixels	1	pixel	=	approximately	1.088 acres
2	"	=	"	1.838	"	2	"	=	"	2.176 "
3	"	=	"	2.757	"	3	"	=	"	3.264 "
4	"	=	"	3.676	"	4	"	=	"	4.352 "
5	"	=	"	4.595	"	5	"	=	"	5.440 "
6	"	=	"	5.514	"	6	"	=	"	6.528 "
7	"	=	"	6.433	"	7	"	=	"	7.616 "
8	"	=	"	7.352	"	8	"	=	"	8.704 "
9	"	=	"	8.271	"	9	"	=	"	9.792 "
10	"	=	"	9.190	"	10	"	=	"	10.880 "

By counting the pixels one can tell how many acres a farm, a field, or an area contain.

Segments, farms, fields, and certain areas can also be located by counting the pixels from left to right on each line. The upper left hand corner of the segment above is located on line 1800 pixel number 2373. "A" Farm area in the upper left hand corner is line 1802, pixel number 2359.